

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/12, C07K 14/47	A2	(11) International Publication Number: WO 99/06548 (43) International Publication Date: 11 February 1999 (11.02.99)
(21) International Application Number: PCT/IB98/01222 (22) International Filing Date: 31 July 1998 (31.07.98) (30) Priority Data: 08/905,135 1 August 1997 (01.08.97) US (71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): DUMAS MILNE ED- WARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire-de-Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). LACROIX, Bruno [FR/FR]; 93, route de Vourles, F-69230 Saint-Genis Laval (FR). (74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Régimbeau, 26, Avenue Kléber, F-75116 Paris (FR).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i> BEST AVAILABLE CO.
(54) Title: 5'ESTs FOR NON TISSUE SPECIFIC SECRETED PROTEINS (57) Abstract The sequences of 5'ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5'ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5'ESTs. The 5'ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5'ESTs. The 5'ESTs may also be used to design expression vectors and secretion vectors.		

CLAIMS

1. A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 38-291 or comprising a sequence complementary thereto.
- 5 2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.
3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-291 or one of the sequences complementary thereto.
4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of
10 one of the sequences of SEQ ID NOs: 38-291 or one of the sequences complementary thereto.
5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.
6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-291 or one of the
15 sequences complementary to the sequences of SEQ ID NOs: 38-291.
7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.
8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-291.
- 20 9. A purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-291 or having a sequence complementary thereto.
10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-291 which encode a signal peptide.
11. A purified or isolated polypeptides comprising a signal peptide encoded by
25 one of the sequences of SEQ ID NOs: 38-291.
12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-291 operably linked to a second nucleic acid encoding a polypeptide.
- 30 13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypeptide into the membrane comprising the steps of:

obtaining a vector according to Claim 12; and

introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

14. A method of importing a polypeptide into a cell comprising contacting said
5 cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-291 operably linked to said polypeptide.

15. A method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-291, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-291;

10 contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-291 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and

isolating said cDNA which hybridizes to said probe.

15 16. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.

17. The cDNA of Claim 16 wherein said cDNA comprises the full protein coding
20 sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

18. A method of making a cDNA comprising one of the sequences of SEQ ID NOs: 38-291, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA;

25 hybridizing said first primer to said polyA tail;

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-291; and

30 isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.

5 20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

21. The method of Claim 18, wherein the second cDNA strand is made by:
contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the
10 sequences of SEQ ID NOs 38-291 and a third primer having a sequence therein which is included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of
15 primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs 38-291, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and
performing a second polymerase chain reaction, thereby generating a second PCR product.

20 22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.

23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding
25 sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

24. The method of Claim 18 wherein the second cDNA strand is made by:
contacting said first cDNA strand with a second primer comprising at least 15
consecutive nucleotides of the sequences of SEQ ID NOs: 38-291;
hybridizing said second primer to said first strand cDNA; and
30 extending said hybridized second primer to generate said second cDNA strand.

25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-291 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.

5 26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

27. A method of making a protein comprising one of the sequences of SEQ ID NO: 292-545, comprising the steps of:

obtaining a cDNA encoding the full protein sequence partially included in one of the
10 sequences of sequence of SEQ ID NO: 38-291;

inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

15 isolating said protein.

28. An isolated protein obtainable by the method of Claim 27.

29. A method of obtaining a promoter DNA comprising the steps of:

obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-291 or the sequences complementary thereto;

20 screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

isolating said DNA comprising said identified promoter.

30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-291 or sequences complementary thereto.

25 31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.

32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

30 33. An isolated promoter obtainable by the method of Claim 32.

34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 292-545.

35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of
5 SEQ ID NOs: 38-291, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-291, or a fragment thereof of at least 15 consecutive nucleotides.

36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of at least 15 consecutive nucleotides.

10 37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of at least 15 consecutive nucleotides.

est

(ix) FEATURE:

(A) NAME/KEY: other
 (B) LOCATION: 93..124
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 96
 region 28..59
 id AA017309
 est

(ix) FEATURE:

(A) NAME/KEY: other
 (B) LOCATION: complement(126..250)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 100
 region 1..125
 id T52392
 est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
 (B) LOCATION: 21..200
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 4.8
 seq LVILSLKSQTLDA/ET

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

```

AGTAAGTCCC CCCGCTCGC ATG ATG GCT GCG GTG CCG CCG GGC CTG GAG CCG   53
      Met Met Ala Ala Val Pro Pro Gly Leu Glu Pro
      -60                      -55                      -50

TGG AAC CGT GTG AGA ATC CCT AAG GCG GGG AAC CGC AGC GCA GTG ACA   101
Trp Asn Arg Val Arg Ile Pro Lys Ala Gly Asn Arg Ser Ala Val Thr
      -45                      -40                      -35

GTG CAG AAC CCC GGC GCG GCC CTT GAC CTT TGC ATT GCA GCT GTA ATT   149
Val Gln Asn Pro Gly Ala Ala Leu Asp Leu Cys Ile Ala Ala Val Ile
      -30                      -25                      -20

AAA GAA TGC CAT CTC GTC ATA CTG TCG CTG AAG AGC CAA ACC TTA GAT   197
Lys Glu Cys His Leu Val Ile Leu Ser Leu Lys Ser Gln Thr Leu Asp
      -15                      -10                      -5

GCA GAA ACA GAT GTG TTA TGT GCA GTC CTT TAC AGC AAT CAC AAC AGA   245
Ala Glu Thr Asp Val Leu Cys Ala Val Leu Tyr Ser Asn His Asn Arg
      1                      5                      10                      15

ATG GGC CGC CAC AAA CCC CAT TTG GCC CTC AAA CAG GTT GAG CAA TGT   293
Met Gly Arg His Lys Pro His Leu Ala Leu Lys Gln Val Glu Gln Cys
      20                      25                      30

TTA AAG CGT TTG ARA AAC ATG AAT TTG GAG GGC GGG   329
Leu Lys Arg Leu Xaa Asn Met Asn Leu Glu Gly Gly
      35                      40

```

(2) INFORMATION FOR SEQ ID NO: 227:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 385 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 39..385
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97
region 1..347
id AA023764
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 146..385
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95
region 145..384
id C03036
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 11..80
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98
region 2..71
id C03036
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 39..231
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99
region 1..193
id R08519
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 232..302
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94
region 193..263
id R08519
est

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 11..109

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 4.8

seq SLVHLLCQNQVLG/NP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

```

AAGTGGCAAG ATG GCG TCC CTG GAT CGG GTG AAG GTA CTG GTG TTG GGA      49
  Met Ala Ser Leu Asp Arg Val Lys Val Leu Val Leu Gly
                -30                      -25

GAC TCA GGT GTT GGG AAA TCT TCG TTA GTC CAT CTC CTA TGC CAA AAT      97
  Asp Ser Gly Val Gly Lys Ser Ser Leu Val His Leu Leu Cys Gln Asn
-20                -15                      -10                      -5

CAA GTG CTG GGA AAT CCA TCA TGG ACT GTG GGC TGC TCA GTG GAT GTC     145
  Gln Val Leu Gly Asn Pro Ser Trp Thr Val Gly Cys Ser Val Asp Val
                1                      5                      10

AGA GTK CAT GAT TAC AAA GAA GGA ACC CCA GAA GAG AAG ACC TAC TAC     193
  Arg Val His Asp Tyr Lys Glu Gly Thr Pro Glu Glu Lys Thr Tyr Tyr
                15                      20                      25

ATA GAA TTA TGG GAT GTT GGA GGC TCT GTG GGC AGT GCC AGC AGC GTG     241
  Ile Glu Leu Trp Asp Val Gly Gly Ser Val Gly Ser Ala Ser Ser Val
                30                      35                      40

AAA AGC ACA AGA GCA GTA TTC TAC AAC TCC GTA AAT GGT ATT ATW NYC     289
  Lys Ser Thr Arg Ala Val Phe Tyr Asn Ser Val Asn Gly Ile Ile Xaa
                45                      50                      55                      60

GTA CAC GAC TTA ACV SAT GGG AAG TCC TCC CAA AAM TTG CGN CGT TGG     337
  Val His Asp Leu Thr Xaa Gly Lys Ser Ser Gln Xaa Leu Arg Arg Trp
                65                      70                      75

TCA TTG GAA GCT CTC AAC AGG GAT TTG GTG CCA ACT GGA GTC TTG GTG     385
  Ser Leu Glu Ala Leu Asn Arg Asp Leu Val Pro Thr Gly Val Leu Val
                80                      85                      90

```

(2) INFORMATION FOR SEQ ID NO: 228:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Brain

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 30..237

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 96
region 12..219

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

```

Met Met Ala Ala Val Pro Pro Gly Leu Glu Pro Trp Asn Arg Val Arg
-60                -55                -50                -45

Ile Pro Lys Ala Gly Asn Arg Ser Ala Val Thr Val Gln Asn Pro Gly
      -40                -35                -30

Ala Ala Leu Asp Leu Cys Ile Ala Ala Val Ile Lys Glu Cys His Leu
      -25                -20                -15

Val Ile Leu Ser Leu Lys Ser Gln Thr Leu Asp Ala Glu Thr Asp Val
      -10                -5                1

Leu Cys Ala Val Leu Tyr Ser Asn His Asn Arg Met Gly Arg His Lys
  5                10                15                20

Pro His Leu Ala Leu Lys Gln Val Glu Gln Cys Leu Lys Arg Leu Xaa
      25                30                35

Asn Met Asn Leu Glu Gly Gly
      40

```

(2) INFORMATION FOR SEQ ID NO: 481:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 125 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (E) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: -33..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8
seq SLVHLLCQNQVLG/NP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

```

Met Ala Ser Leu Asp Arg Val Lys Val Leu Val Leu Gly Asp Ser Gly
      -30                -25                -20

Val Gly Lys Ser Ser Leu Val His Leu Leu Cys Gln Asn Gln Val Leu
      -15                -10                -5

Gly Asn Pro Ser Trp Thr Val Gly Cys Ser Val Asp Val Arg Val His
  1                5                10                15

Asp Tyr Lys Glu Gly Thr Pro Glu Glu Lys Thr Tyr Tyr Ile Glu Leu
      20                25                30

```

Trp Asp Val Gly Gly Ser Val Gly Ser Ala Ser Ser Val Lys Ser Thr
 35 40 45

Arg Ala Val Phe Tyr Asn Ser Val Asn Gly Ile Ile Xaa Val His Asp
 50 55 60

Leu Thr Xaa Gly Lys Ser Ser Gln Xaa Leu Arg Arg Trp Ser Leu Glu
 65 70 75

Ala Leu Asn Arg Asp Leu Val Pro Thr Gly Val Leu Val
 80 85 90

(2) INFORMATION FOR SEQ ID NO: 482:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: -31..-1
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.8
seq WAFSCGTWLPSRA/EW

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

Met Val Phe Pro Ala Lys Arg Phe Cys Leu Val Pro Ser Met Glu Gly
 -30 -25 -20

Val Arg Trp Ala Phe Ser Cys Gly Thr Trp Leu Pro Ser Arg Ala Glu
 -15 -10 -5 1

Trp Leu Leu Xaa Val Arg Ser Ile Gln Pro Glu Glu Lys Glu Arg Ile
 5 10 15

Gly Gln Phe Val Phe Ala Arg Asp Ala Lys Ala Ala Met Ala Gly Arg
 20 25 30

Leu Met Ile Arg Lys Leu Val Ala Glu Asn Arg
 35 40

(2) INFORMATION FOR SEQ ID NO: 483:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☒ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.